

On Solving Nonsmooth Retail Portfolio Maximization Problems Using Active Signature Methods

Aswin Kannan*, Timo Kreimeier, and Andrea Walther

December 22, 2023

Abstract

The retail industry is governed by crucial decisions on inventory management, discount offers like promotions and stock clearing as so-called markdowns, presenting two sets of optimization problems. The former is an estimation problem, where the underlying objective is to predict the coefficients of demand (sales) elasticity with respect to product prices. The latter is the dynamic revenue maximization problem, which takes in the coefficients of demand as inputs. While both tasks present nonsmooth optimization problems, the latter is a challenging nonlinear problem in massive dimensions. This is further subject to constraints on inventory, inter-product relationships, and price bounds. Traditional approaches to solve such problems relied on using reformulations and approximations, thereby leading to potentially suboptimal solutions. In this work, we retain the nonsmooth structure generated by the max type (or equivalently absolute value type) function and solve the resulting problem in its abs-quadratic form, i.e., in a quadratic matrix-vector-product based representation including linear arguments in the abs-evaluation. Subsequently, we present an adaptation of the Constrained Active Signature Method (CASM) that explicitly exploits this abs-quadratic structure of the problem yielding the Quadratic Constrained Active Signature Method (QCASM). In process, we also guarantee convexity of the objectives under some mild realistic assumptions on the market demand and structure. Two real world retail examples (UK and US market data from 2017-2019) and one simulated use-case are studied from an empirical standpoint. Numerical results demonstrate good performance of QCASM and further show that such solvers can be used significantly by the retail science community in the future.

1 Introduction and Survey

Retail markets primarily gravitate towards revenue maximization. While the revenue can simply be expressed as a direct product of price and sales, these two quantities are actually more complicated in reality [Coh+17; CPP16; KPW20] as they depend on multiple factors like consumer demand, time-line, inventory levels, and relationships between SKUs. Here, SKU refers to Stock Keeping Unit, or in simpler terms product ID. It can also be noted that pricing and consumer demand vary significantly from one SKU to another, e.g., pricing of a T-shirt may differ significantly from that of rice or sugar. Consumer demand is usually a consequence of prices, social aspects, geographical indicators, and many other criteria [All38; CDH90]. Retailers also face challenges in maintaining sufficient inventory levels at their respective outlets that are subject to several logistic factors, practical weather aspects, and storage limitations. Within the same SKU, the demand towards different sizes can vary. For example, the demand for a medium size T-shirt may be higher than that of a large or small T-shirt. Proposing a different SKU to each such size is deemed to be arduous and less robust, since this can significantly increase the problem dimension. One way to handle such variations is by deploying the stock response factor [Li+22], which models demand as a function of both customer preferences and available inventory levels.

Another crucial aspect with retail markets is the aspect of discounts [Coh+17; CPP16; KP20; KPW20]. Discounts are usually offered in two levels, respectively referred to as promotions and markdowns. Promotions are lower levels of discounts intended to increase product sales temporarily. These reductions are usually enforced when a new competitive product is introduced or if sudden changes in inventory are required. Products are also promoted when there is a sudden random surge/drop in customer demand levels. Markdowns are slightly different and are offered during inventory clearance. These prices are considerably lower than the base and are brought forward as a mechanism to either discard older stocks or when replacing the SKU, e.g., when terminating a product from the catalogue. There exist constraints on

*All the authors are from Humboldt-Universität zu Berlin, Unter den Linden 6, 10117 Berlin, Germany. The corresponding author is reachable at “aswin.kannan@hu-berlin.de”.

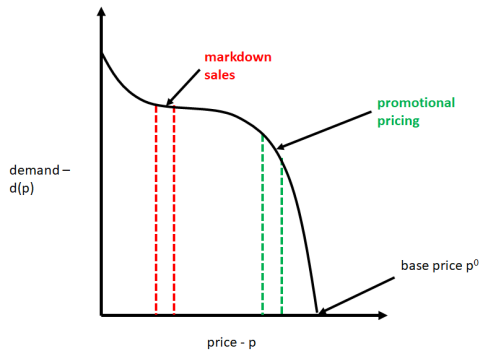


Figure 1: Demand-price curve for regular products/SKUs.

the number of times promotions or markdowns are offered, since they are tied to several internal policies and regulations of the corresponding businesses. Additionally, there also exist inter-item constraints within multiple SKUs, e.g., SKU “A” cannot be promoted when SKU “B” is promoted. Lastly, the entire problem presents a dynamic programming framework, where decision variables are coupled in time.

It is practically obvious and well established that consumer demand is monotonically decreasing in terms of price, predominantly following an S-like pattern as shown in figure 1, [All38, chapter 5], or [CPP16]. Here, vaguely the region between the markdowns and the base price is concave. The overall curve can be neither convex or concave and the patterns are not the same as above for all SKUs. Usually, the windows of promotions and markdowns respectively are between 70-80 percent and 30-50 percent of the base price. However, this is not rigid and the window can be wider/shorter depending on the product type. Quite often, products are priced between the base price and promotional levels. From the standpoint of analysis, both convex [HLP13; SRL08] and concave [CRS06; DHP10; HLP13] instances of demand-price curves have been studied in detail. Let $d_i^t(p_i^t)$ denote the demand of product i at time t depending on the price p_i^t . Then, some examples of demand functions are given by

$$\underbrace{d_i^t(p_i^t) = a_i - b_i (p_i^t)^\gamma}_{\text{concave for } \gamma \geq 0}, \quad \underbrace{d_i^t(p_i^t) = (a_i - b_i p_i^t)^\gamma}_{\text{concave for } \gamma \leq 1 \text{ and convex for } \gamma > 1}, \quad \underbrace{d_i^t(p_i^t) = a_i (p_i^t)^{-b_i}}_{\text{convex}}, \quad (1)$$

where a_i , b_i , and γ are positive scalars. Comprehensive details on different choices of demand-pricing functions have been discussed in [HLP13]. Some works attempt to address the dynamic aspects of demand with respect to price [Coh15; Coh+17] using for example functions like

$$d_i^t(p_i^t) = (p_i^t)^{\alpha_1} (p_i^{t-1})^{\alpha_2} \dots (p_i^{t-K})^{\alpha_{K+1}}.$$

In our work, we do not focus on this aspect and assume that the demand versus price curve is static. The structure of revenue is given by the function $p_i^t d_i^t(p_i^t)$, which can be a highly nonlinear function in p_i^t . Additionally, this optimization problem is subject to several constraints on promotions and inventory levels.

Underlying Nonsmoothness Besides a few practical implementations, literature has mainly addressed these problems from a discrete perspective [Coh+17; Hua+21]. The resulting optimization tasks are either solved using standard mixed-integer linear programming solvers like CPLEX or Gurobi for smaller dimensions [Coh+17; SS10] or solved using heuristics-based approaches for larger dimensional instances [Hua+21; MF17]. There have been some works based on a continuous problem formulation [YGT14]. We focus on solving the continuous version of this problem, allowing for prices to take any intermediate value between the lower and upper bounds. Closed form expressions for demand are usually unknown and are highly SKU specific. Estimation of these demands has been primarily based on data. Some approaches have attempted to use neural networks or decision trees to come up with such estimations [Mit+22], when the demand is a function of various product design attributes in addition to price. This is beyond the scope of our work, where the focus is only on prices. Since nonlinear demand functions have their own issues with tractability, we concentrate on solving piecewise linear instances of demand in our work, covering almost all aspects of retail markets, as we will see later through the paper. The following instances are examples where nonsmoothness arises in retail demand.

- Given historical data, it has been observed that piecewise linear functions prove to be robust in quantifying the demand-price behavior, [Coh15, chapter 4], or [KP20; KPW20]. In cases of convex

and concave demand, their respective approximations take the following form.

$$d_i^t(p_i^t) \approx \max(\tilde{a}_i - \tilde{b}_i p_i^t, a_i - b_i p_i^t), \quad \text{or} \quad d_i^t(p_i^t) \approx \min(\tilde{a}_i - \tilde{b}_i p_i^t, a_i - b_i p_i^t).$$

It can be observed that more than two pieces can be deployed in the above expressions without loss of generality.

- Demand-price models are inherently piecewise linear in the presence of reference price effects [Che+16] and non-negativity of demand.

$$d_i^t(p_i^t) = \underbrace{\left\{ \begin{array}{ll} a_i - b_i p_i^t, & p_i^t \leq \bar{p}_i \\ a_i - b_i p_i^t - c_i(p_i^t - \bar{p}_i), & p_i^t > \bar{p}_i. \end{array} \right\}}_{\text{reference price effects}}, \quad \underbrace{d_i^t(p_i^t) = \max(a_i - b_i p_i^t, 0)}_{\text{demand non-negativity}}.$$

Demand non-negativity can also be written in a form similar to the expression for reference price effects, when $\bar{p}_i = \frac{a_i}{b_i}$, and $c_i = 1$.

- The stock response factor (mentioned earlier) has been noted to be a highly nonlinear function. Approximating the same using piecewise linear functions has also been well accepted as stated in [Li+22]. This can be expressed as follows.

$$d_i^t(p_i^t, I_i^t) = \text{sr}f(I_i^t) \cdot r_p(p_i^t), \quad \text{sr}f(I_i^t) = r_s(I_i^t) \approx \max(\tilde{a}_i - \tilde{b}_i I_i^t, a_i - b_i I_i^t),$$

where I_i^t refers to the inventory of SKU i at time t , r_p to the function corresponding to seasonality and r_s to the actual stock response factor.

For this work, we consider prices to be continuous variables and demand to be a consequence of piecewise linear functions.

Active Signature Methods A natural question that would arise is the possible use of slack variables (sales) s_i^t to reformulate the problem into a smoother version by deploying constraints and incorporating terms of the form $p_i^t s_i^t$ in the objective. For example,

$$d_i^t(p_i^t) = \min(\tilde{a}_i - \tilde{b}_i p_i^t, a_i - b_i p_i^t)$$

can be presented as constraints of the form

$$\left\{ s_i^t \leq \tilde{a}_i - \tilde{b}_i p_i^t, s_i^t \leq a_i - b_i p_i^t, \quad \forall i, t \right\}.$$

While such reformulations work reasonably well in practice, we note that these do not present guarantees on global optimality and can potentially lead to suboptimal solutions, unless computationally expensive global solvers are used. In cases where the demand function $d_i^t(p_i^t)$ is concave, convexity of the objectives is preserved when using the primitive formulation. These optimization problems fall under the category of ‘‘Extended Quadratic Linear Programs’’ (EQLPs). We propose an appropriate adaptation of the active signature method [GW18; KWG21] as alternative to solve our EQLP. While [KWG21] focused on solving a problem with piecewise linear objectives, the novelty in this work arises from extending the same to quadratic objectives of a certain structure. We name our solver Quadratic Constrained Active Signature Method (QCASM). More details of this method are explained in section 3.

Contributions: As parts of the larger revenue maximization problem, we focus on two different questions.

- Prediction: Given historical data on price and demand for each SKU, we solve a regression problem to obtain our nonsmooth demand functions. This problem is a piecewise linear optimization problem and takes the form

$$\min_{a, b \geq 0} \sum_{i \in \mathcal{I}} \sum_{j=1}^n \left| f_i(a, b, p) - d_j^{obs} \right|,$$

where $f_i(a, b, p)$ is a PLF (piecewise linear function) of the min or max type. We deploy the extended active signature method mentioned above to solve the same. We consider three types of PLF in this paper. They are defined as follows:

$$\underbrace{\max(\tilde{a} - \tilde{b}p, a - bp)}_{\text{Convex PLF-1}}, \quad \underbrace{\max(\tilde{a} - \tilde{b}p, 0)}_{\text{Convex PLF-2}}, \quad \underbrace{\min(\tilde{a} - \tilde{b}p, a - bp)}_{\text{Concave PLF}}.$$

- Revenue maximization: Setting the demand, $d_i^t(p_i^t) = f_i(a, b, p_i^t)$, we deploy QCASM to solve the larger problem of interest, i.e., the revenue maximization problem, that takes the form

$$\min_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{F}} -f_i(a, b, p_i^t) p_i^t.$$

Note that the constraint set \mathcal{P} and index sets \mathcal{T}, \mathcal{F} are defined appropriately (more details in the next section).

While literature has analyzed practical datasets pertaining to this problem, they are not publicly available [Bur+21; Hek+19]. There exist some open source datasets [Che; SB21; Ske+21] that present similar data, but they are tailored towards clustering applications and product attributes other than price. Hence, we resort to three widely used open source datasets [Coh+22; CSG12; EC23] for our studies, which explicitly contain the price-sales data for multiple years and SKUs for individual retailers, as detailed in section 4. The remainder of the paper is organized into four sections. Section 2 presents the mathematical model. Section 3 discusses the extended active signature method in detail. Section 4 demonstrates the behavior of QCASM on three practical datasets and we conclude in section 5.

2 Problem Formulation

Our problem formulation centers on pricing and inventory management. We consider constraints on inventory, bounds on promotions and markdowns, coupled item pricing, and replenishments. At this point, we once again mention that our focus is on the continuous version of the problem. Along the lines of earlier works [Coh+17; KPW20], the revenue maximization problem can be put forth as follows.

$$\begin{aligned} \min_{p, u, I} h(p) &= \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \left(\underbrace{-p_i^t d_i^t(p_i^t)}_{\text{revenue}} + \underbrace{\phi(d_i^t(p_i^t))}_{\text{costs}} \right) & (2) \\ \text{subject to } I_i^{t+1} &= I_i^t + u_i^t - d_i^t(p_i^t) & \text{Inventory Dynamics constraint} \\ \rho_i^l &\leq p_i^t \leq \rho_i^u, \quad \forall t \in \mathcal{T}_p & \text{Promotion constraint} \\ \theta_i^l &\leq p_i^t \leq \theta_i^u, \quad \forall t \in \mathcal{T}_m & \text{Markdown constraint} \\ p_i^t - p_j^t &\geq \kappa_{ij}, \quad \forall \{i, j\} \in \mathcal{N} & \text{Inter-Item constraints} \\ u_i^t &= 0, \quad \forall t \in \mathcal{T}_i & \text{Non-replenishment time-slots} \\ I_i^t, u_i^t, p_i^t &\geq 0, \quad \forall i \in \mathcal{N}, t \in \mathcal{T} & \text{Non-negativity constraints.} \end{aligned}$$

Here, I_i^t , p_i^t , and u_i^t respectively refer to the inventory, price, and replenishments of product i at time t . During several times of the year, stores do not receive any replenishments due to regulations, transportation issues, and sudden power outages. Discounts, i.e., promotions and markdowns, are mostly seasonal and are offered at specific sets of time \mathcal{T}_p and \mathcal{T}_m . An example of inter-item constraints is the following: If milk is priced at 2 Euros per Litre, then yogurt should be priced at the least around 1.5 Euros per Litre. The demand $d_i^t(p_i^t)$ predominantly follows different trends and can turn out to be discontinuous, nonlinear, and nonconvex. Some real world examples of historical data from a retail store are shown in figure 2. Given that there are no closed form expressions, we attempt to approximate the demand by means of piecewise linear functions. There has been extensive research along these lines for several types of demand functions [HLP13]. Recent granted patent works [KP20] use one such approximation to solve a discrete version of our problem, where $d_i^t(p_i^t) \approx \max_j (h_j(p_i^t))$, $h_j(p_i^t) = a_i^j + b_i^j p_i^t$.

2.1 Demand Types

While there are innumerable ways to approximate nonlinear demand by piecewise linear functions, we focus on the following three cases. While this is not exhaustive, it still covers most of the pricing patterns in the retail industry.

2.1.1 Convex Demand

An exhaustive survey of convex models has been described in [HLP13]. This type of behavior stems in both competitive (oligopolistic) and single-firm based settings. In the case of competition, it has been well accepted that demand either falls under the category of discrete choice or Multiplicative Competitive Interaction (MCI) models. Discrete choice models are either logistic [KC14] or probit type [Chi01], which are majorly concave. As an alternative to logit models, the Multiplicative Competitive Interaction (MCI) model [KC14; NC74] has also been studied in great detail. A static example (dropping time suffixes) of

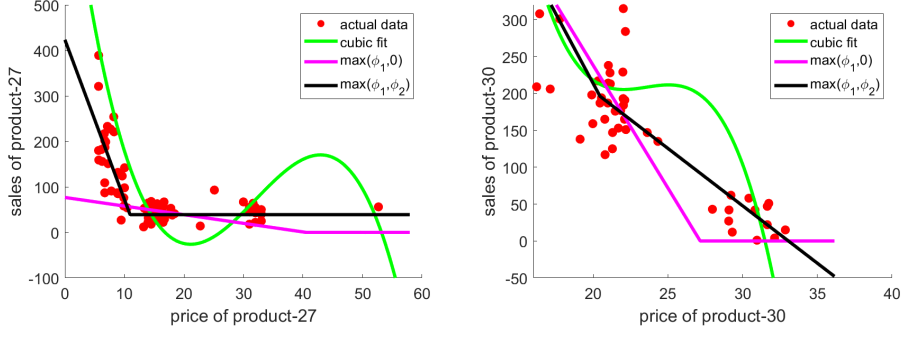


Figure 2: Examples of demand-price curves for regular products from [Coh+22].

MCI type models in lieu of competition is given in Equation (3). Here, p_{-i} refers to prices of agents (retailers) other than i . Logarithmic type models (log-log) of the form given by Equation (3) have also been studied in literature [Coh+17]. As mentioned earlier, [Coh+17] also presents an MCI model in the context of dynamic and non-competitive regimes. Some sample data profiles (historical sales-price data) that depict convex behavior have been shown already in figure 2.

$$\text{MCI: } d_i(p_i, p_{-i}) = a_i p_i^{b_i} \prod_{j \neq i} p_j^{b_j}, \quad \text{Log-Log: } \log(d_i(p_i^t)) = a_i \log(p_i^t) + b_i. \quad (3)$$

We further highlight a few important model classes from literature as follows.

- Linear type: These take a simple form $d(p) = a - bp$ and have been extensively studied [Mil59]. Nonsmooth counterparts related to price caps and reference prices have also been studied (to be discussed later).
- Iso-elastic type: These have a sharper elasticity or response towards price [KC62] and have also been widely studied in literature. These are of the form $d(p) = ap^{-b}$.
- Power Linear type: These are inverse demand functions that are represented as follows [SRL08]: $d(p) = (a - bp)^\gamma$, $\gamma > 1$.
- Hybrid versions [LL03]: They are combinations of any of the above three forms of functions (additive or composition).

Smooth and convex demand functions can be approximated by a max over multiple linear functions, i.e.,

$$d(p) = \max_i (a_i + b_i p), \quad a_i, b_i \geq 0.$$

One of the main disadvantages with this setting is that the revenue function, which is a product of negative price times demand happens to be nonconvex. In this relation, we present the following theorem on convexity. We note that the related extensions follow by direct inspection.

Theorem 2.1. *Let $f(p) = -p \cdot \max(a - bp, 0)$ and let $a, b \geq 0$. Then, the function $f(p)$ is not convex over the complete domain $-\infty \leq p \leq \infty$.*

Proof. Let us consider p_1 and p_2 such that $a - bp_1 > 0$ and $a - bp_2 < 0$. That is, $p_2 > \frac{a}{b} > p_1$. Let $p_3 = \alpha p_1 + (1 - \alpha)p_2$ for $\alpha \in (0, 1)$. It is easy to observe that $p_2 > p_1$ and $f(p_1) < f(p_2) = 0$. This further implies

$$\alpha f(p_1) + (1 - \alpha)f(p_2) = \alpha f(p_1) < 0. \quad (4)$$

Noting the strict inequalities, it is easy to observe that there exists $\bar{\alpha} \in (0, 1)$ such that

$$p_3 = \bar{\alpha} p_1 + (1 - \bar{\alpha})p_2, \quad \text{where } p_3 = \frac{a}{b}.$$

Using Equation (4), this implies that $f(p_3) = f(\alpha p_1 + (1 - \alpha)p_2) = 0 > \alpha f(p_1) + (1 - \alpha)f(p_2)$. This completes the proof. ■

Remark: For max over three functions and higher, recursive arguments can be made by using expressions similar to $\max(f_1, f_2, f_3) = \max(\max(f_1, f_2), f_3)$.

Stock Response Factor: This setting is very common in domains pertaining to fashion. Here, the differences between sales and available inventory can be enormous [CG05] and only a subset of products are kept on display. As an example, consider a firm that may have 100 white T-shirts in their stock. The size is say, spread over “Small”, “Medium”, “Large”, and “XL”. The distribution may not be exactly 25 each. The display may start, e.g., with 10 T-shirts per size. During the course of the day this could be reduced to one T-shirt of size “Small”. This may not exactly correspond to the situation in the inventory. Usually there are minor price differences between sizes among the same brand, but these are assumed to follow fixed patterns. Here, the sales capacity is limited by the following expression

$$s_i^t \leq \text{srf}_i^t(I_i^t) \cdot \underbrace{d_i^t(p_i^t)}_{\text{constant}}.$$

Since the function “srf” is very nonlinear, the well accepted methodology in literature is to approximate it by a set of piecewise linear functions [Li+22] as follows: $\text{srf}_i^t(I_i^t) = f_{\text{nonlin}}(I_i^t) \approx \max_j(b_j I_i^t + a_j)$, $a_j, b_j \geq 0$. Here, the objective is to maximize sales and prices are assumed to be fixed as per seasonal rules within a firm. Note that the above inequality automatically presents a nonconvex constraint in s and I . We do not focus on “srf” in this work, but mention it in this subsection, because it follows a similar problem structure as our case of convex demand.

2.1.2 Nonconvex demand

Literature’s focus has also been significant on concave demand type models. These may be classified vaguely into three forms as follows. This classification is not existent in literature and we merely state it for the ease of understanding.

- Logit models: These choice models as stated previously in [Chi01; KC14] have predominantly been the mainstay of literature. Similar to price p , we note that product attributes a_p like design specification become very relevant to the question. An example of a logit model is given as follows.

$$d(p, a_p) = \eta_p \left(1 + e^{-U(p, a_p)}\right)^{-1}.$$

The utility function $U(\cdot)$ is usually linear or quadratic. Consider a winter-wear product (shawl or gloves). Let p and a_1 denote the price and number of woolen layers respectively. Then, the utility function may take the form $U(p, a_p) = 0.1p - 0.01a_1$. The equation above is a case restricted to a single-firm with the customer presented with two options, namely purchase (high utility) and no-purchase (the scalar term “1” indicates the no-purchase option or no-utility). The purchase levels are fractions or probability levels, which are in turn multiplied by the market population η_p . We note that this extends to a multinomial-logit model [CDH90] when there are multiple firms, each with an exponential component defined in the denominator. However, we do not focus on product attributes in this work.

- Exponential models [CSL04]: These models can be defined by $d(p) = a - be^{\gamma p}$, where $a, b, \gamma \geq 0$ are scalars. Exponential models are structurally slightly different from logit models.
- Logarithmic models [CRS06; RLS05]: Their structure is also similar to exponential models and they take the form: $d(p) = (\log a - bp)^\gamma$.
- Polynomial models [SRL08]: These models are more commonly deployed in several industries and are given by $d(p) = (a - bp)^\gamma$ for $\gamma < 1$. Above, we had discussed cases where these models are convex for $\gamma > 1$. Concave demand models of the form $d(p) = a - bp^\gamma$ have also been studied [CRS06].

Some examples of data depicting nonconvex demand functions are given in figure 3. We note that when demand is concave, then the resulting minimization problem happens to be convex. We present the formal statement as follows.

Theorem 2.2. *Let $f(p) = -p \cdot \min(\tilde{a} - \tilde{b}p, a - bp)$ and let $\tilde{a}, \tilde{b}, a, b \geq 0$. Then, the function $f(p)$ is convex over the complete domain $-\infty \leq p \leq \infty$.*

Proof. Using a similar argument along the lines of theorem 2.1, this result follows. The proof can also be taken from theorem 2.4, which is discussed later in this section. ■

2.1.3 Relativistic Demand Pricing

Some retail portfolios are characterized by relative pricing, where the drop in demand becomes significantly higher when the price crosses a threshold. This is very commonly studied in both game-theoretic regimes [FGL03] and single-agent optimization settings [Gre95]. Such a structure has also received considerable attention in marketing-based retail settings with multiple channels [CCH03]. Retail industries

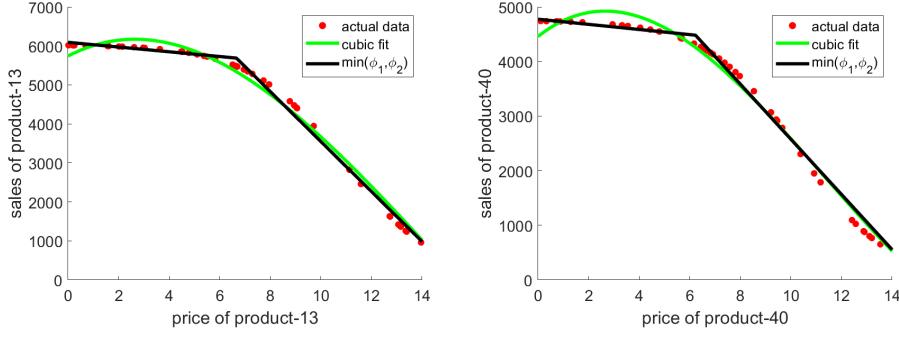


Figure 3: Concave example of a demand-price curve for products for the logit demand from [EC23].

may have both physical store based outlets and online stores. While the indirect channels do not have many reference price effects, direct channels have shown a strong requirement for such nonsmooth demand behavior as follows.

$$d^t(p^t) = a - bp^t - \gamma(p^t - p_{\text{ref}}), \quad \gamma = \begin{cases} \gamma_{\text{ref}}, & p^t > p_{\text{ref}} \\ 0, & p^t \leq p_{\text{ref}} \end{cases}.$$

Positive sales: In simpler cases, where the demand function is linear and smooth, it can still be noticed that the sales have to be non-negative [KC14]. This is dictated by

$$s^t = \max(d^t, 0) = \max(a - bp^t, 0).$$

A natural question that would arise is whether this behavior can be simply invoked in the problem by plugging in s^t instead of d^t in the objective and imposing constraints $s^t \geq 0$ and $s^t \geq d^t$. However, such a formulation induces bilinear terms of the form $p^t s^t$ in the objective, and we lose the benefits of the primitive formulation as discussed earlier. Ignoring the domain of prices, where demand can be negative, is not practical from a business perspective.

Zero price-elasticity: In wholesale outlets, demand responses can be significantly low or even zero for certain levels of price. For example, assume that the regular price of a milk is 1.2 Euros per litre and the demand at this level is 1000 units. If the price drops to 0.5 Euros, the demand increases to 5000 units. If the price drops further down to 0.2 Euros, the demand still stays at 5000 units. Beyond a point, the society becomes “in-elastic” to price changes. This effect can be described by the expression $d(p) = \min(a - bp, \bar{d})$. This setting is also analogous to price-caps in Cournot based regimes [TR19], where prices are defined as a consequence of demand, e.g., $p(q) = \min(a - bq, \bar{p})$. In both of these cases, the revenue functions lead to convex optimization problems (similar structure to the functions in theorem 2.4). We now move onto our learning problem, i.e., the fitting of piecewise linear functions to data.

2.2 Regression Subproblem

It can be observed that building “parametric” models based on data automatically leads to regression based settings. Given that we already have piecewise linear functions in our model to be estimated, we move along the path of l1-regression, which further presents piecewise linear functions in the absolute value form, i.e., using $|\cdot|$. The benefit of such an l1-type setting is that the learning problem yields an optimization formulation with piecewise linear objectives, which can be handled by our QCASM framework. Moving to l2-type regression (or polynomial and more complex models) further makes the problem nonlinear, obviously necessitating the use of more stylized solvers and reformulations. We do not intend to compare the complexity versus accuracy behavior of multiple regression models in this work and leave this large area for future research. Noting that all the three types of PLFs lead to similar formulations, we consider the regression problem for PLFs of the type $d_i^t(p_i^t) \approx \min(\tilde{a}_i - \tilde{b}_i p_i^t, a_i - b_i p_i^t)$ in the following form:

$$\begin{aligned} \min_{a, \tilde{a}, b, \tilde{b}} \quad & \sum_{i \in \mathcal{I}} \sum_{k=1}^K \left| \min(\tilde{a}_i - \tilde{b}_i p_i^{k,sl}, a_i - b_i p_i^{k,sl}) - d_i^{k,sl} \right| \\ \text{subject to} \quad & \tilde{a}_i, a_i, \tilde{b}_i, b_i \geq 0 \quad \forall i \in \mathcal{I}. \end{aligned} \quad (5)$$

Here, \tilde{a}_i, a_i and \tilde{b}_i, b_i respectively refer to the intercept and slope of the demand function and

$$\left\{ p_i^{k,sl}, d^{k,sl} \right\} \quad 1 \leq k \leq K$$

to the price-sales data, i.e., the history from retailer records. We note that the regression problem also need not necessarily be convex by means of the following corollary.

Corollary 2.3. *Let $h(a, b) = |\max(a - bp, 0) - d|$ for scalars $p, d \geq 0$. Then, $\min_{a, b \geq 0} h(a, b)$ does not necessarily present a convex problem.*

Proof. Using the same argument as in theorem 2.1, this result follows. The steps are straightforward. \blacksquare

We move to the next subsection to analyze the formulation of the larger revenue maximization problem from the angle of a composition of absolute value type functions. As we will see later through the next section, this fits into the general framework of EQLPs, for which our solver QCASM can be deployed.

2.3 Larger problem in abs-quadratic form

We state our formulation for the case (2) as minimum over two functions. Note that the flipping of signs (maximum over two functions) and subtraction of linear terms (maximum over a linear function and 0) lead to the other two types of PLFs.

$$\begin{aligned} \min_{p, I} f(p) &= - \sum_{i \in \mathcal{I}} \sum_{t=0}^{T-1} p_i^t \min \left(\tilde{a}_i - \tilde{b}_i p_i^t, a_i - b_i p_i^t \right) = \sum_{i \in \mathcal{I}} \sum_{t=0}^{T-1} p_i^t \max \left(\tilde{b}_i p_i^t - \tilde{a}_i, b_i p_i^t - a_i \right) \\ \text{subject to : } I_i^{t+1} &= I_i^t + r_i^t + \max \left(b_i p_i^t - a_i^0, b_i p_i^t - a_i \right) \\ p_{i,l} &\leq p_i^t \leq p_{i,u}, \quad \forall t \in \mathcal{T} && \text{Bound constraints on price} \\ I_i^t &\geq 0, \quad \forall i \in \mathcal{I}, t \in \mathcal{T} && \text{Non-negativity constraints} \\ I_i^T &\leq \bar{I}, \quad \forall i \in \mathcal{I} && \text{Final-inventory constraint.} \end{aligned}$$

The promotions and markdown constraints are jointly incorporated as bound constraints in the above expression. Introducing additional variables, the problem can be rewritten in the following form:

$$\begin{aligned} \min_{I, p, s, \hat{s}} \hat{f}(p, s, \hat{s}) &= \sum_{i \in \mathcal{I}} \sum_{t=0}^{T-1} p_i^t (b_i p_i^t - a_i) + \frac{1}{2} \sum_{i \in \mathcal{I}} \sum_{t=0}^{T-1} p_i^t (s_i^t + \hat{s}_i^t) \\ \text{subject to } s_i^t &= (b_i^0 - b_i) p_i^t - a_i^0 + a_i, \quad \forall t \in \mathcal{T} && \text{(EQ-1)} \\ 2I_i^{t+1} &= 2I_i^t + 2r_i^t + 2b_i p_i^t - 2a_i + s_i^t + \hat{s}_i^t, \quad \forall t \in \{0, \dots, T-1\} && \text{(EQ-1)} \\ s_i^t &= a_i - b_i p_i^t, \quad \forall t \in \{0, \dots, T-1\} && \text{(6)} \\ p_{i,l} &\leq p_i^t \leq p_{i,u}, \quad \forall i \in \mathcal{I}, t \in \mathcal{T} && \text{(INQ-1)} \\ I_i^t &\geq 0, \quad \forall i \in \mathcal{I}, t \in \mathcal{T} && \text{(INQ-1)} \\ I_i^T &\leq \bar{I}, \quad \forall i \in \mathcal{I} && \text{(INQ-1)} \\ \hat{s}_i^t &= |s_i^t|, \quad \forall i \in \mathcal{I}, t \in \mathcal{T} && \text{(EQ-2),} \end{aligned}$$

using $\max(s_i^t, 0) = (s_i^t + |s_i^t|)/2$. Here, note that (INQ-1) refers to inequality constraints of a particular form handled by QCASM (described later) and (EQ-1) and (EQ-2) refer to two similar classes of equality constraints. For the ease of parsing and clarity, we maintain the same terminology in both (6) and (CAQOP) discussed in section 3 for constraint classes (EQ-1), (EQ-2), and (INQ-1). Next, we study the convexity of the objective \hat{f} .

Theorem 2.4. *Consider problem (6) and the function \hat{f} . Assume that there exists a feasible point (p, s, \hat{s}) that satisfies all the constraints (EQ and INQ). Then, in the feasible region governed by the constraints (EQ and INQ), the function \hat{f} is convex.*

Proof. Exploiting the structure of \hat{f} , it suffices to prove the convexity of the following function

$$\tilde{f}(p) = 2p(bp - a) + p \left((\tilde{b} - b)p - \tilde{a} + a + |(\tilde{b} - b)p - \tilde{a} + a| \right),$$

for given and fixed values of $\tilde{a}, a, \tilde{b}, b \in \mathbb{R}$. Note that we drop the subscripts of i pertaining to products for the simplicity of notation and also that $p \in \mathbb{R}$ for this discussion. Let $\tilde{b} = b + \delta, \tilde{a} = a + \gamma$. Then, \tilde{f} can be further simplified to

$$\tilde{f}(p) := (2b + \delta)p^2 + |\delta p^2 - \gamma p|.$$

Setting $\lambda = \frac{\gamma}{\beta}$, then it suffices to prove the convexity of the function

$$\tilde{g}(p) := p^2 + |p^2 - \lambda p|, \quad \lambda \geq 0.$$

Let p_A and p_B refer to any two arbitrary prices that satisfy the bounds, i.e., the inequality constraints INQ in (6). Then, it suffices to show that

$$\tilde{g}(\alpha p_A + (1 - \alpha)p_B) \leq \alpha \tilde{g}(p_A) + (1 - \alpha)\tilde{g}(p_B) \quad (7)$$

holds. We prove this by an exhaustive case-by-case analysis. Let $p_C = \alpha p_A + (1 - \alpha)p_B$

- $p_A, p_B \geq \lambda$: Then, Equation (7) follows from the convexity of the quadratic function $(2p^2 - \lambda p)$.
- $p_A, p_B \leq \lambda$: Here, $\tilde{g}(p_A) = \lambda p_A$ and $\tilde{g}(p_B) = \lambda p_B$ holds. Then, the result follows from the convexity of the linear function λp .
- $p_A \leq \lambda \leq p_C \leq p_B$: For the expressions in Equation (7) one obtains

$$\begin{aligned} \tilde{g}(\alpha p_A + (1 - \alpha)p_B) - \alpha \tilde{g}(p_A) + (1 - \alpha)\tilde{g}(p_B) &= 2((\alpha p_A + (1 - \alpha)p_B)^2 - \alpha \lambda p_A - (1 - \alpha)p_B^2) \\ &= 2(\alpha^2 p_A^2 + p_B^2(\alpha^2 - \alpha) + 2\alpha(1 - \alpha)p_A p_B - \alpha \lambda p_A) \\ &= 2(\alpha^2 p_A^2 - p_B^2 \alpha(1 - \alpha) + 2\alpha(1 - \alpha)p_A p_B - \alpha \lambda p_A) \\ &\leq 2(\alpha^2 p_A^2 - p_B^2 \alpha(1 - \alpha) + 2\alpha(1 - \alpha)p_A p_B - \alpha p_A^2) \\ &= 2(-\alpha(1 - \alpha)p_A^2 - \alpha(1 - \alpha)p_B^2 + 2\alpha(1 - \alpha)p_A p_B) \\ &= -2\alpha(1 - \alpha)(p_A - p_B)^2, \leq 0 \end{aligned}$$

proving the assertion for this case.

- $p_A \leq p_C \leq \lambda \leq p_B$: Then one obtains for the expressions in Equation (7) that

$$\begin{aligned} \tilde{g}(\alpha p_A + (1 - \alpha)p_B) - \alpha \tilde{g}(p_A) + (1 - \alpha)\tilde{g}(p_B) &= \lambda(\alpha p_A + (1 - \alpha)p_B) - \alpha \lambda p_A - (1 - \alpha)(2p_B^2 - \lambda p_B) \\ &= -(1 - \alpha)2p_B^2 + 2\lambda(1 - \alpha)p_B = 2(1 - \alpha)p_B(\lambda - p_B), \leq 0 \end{aligned}$$

proving the assertion for this case.

Hence, the convexity of our objective $f(\cdot)$ is shown. \blacksquare

We conclude this section by mentioning that the above theorem on convexity does not hold for the other two instances of PLF given by the $\max(\tilde{a} - \tilde{b}p, a - bp)$ and $\max(a - bp, 0)$ types.

3 Optimization Approach

In this section, we will present a solution algorithm for the optimization problem posed previously. Since this algorithm is not specific and exclusive to the problem in this paper and builds on a previously published algorithm, the notation in this section is detached from the other sections.

For solving unconstrained piecewise linear optimization problems, Andreas Griewank and Andrea Walther presented the Active Signature Method (ASM) in [GW18]. The basic idea of this method is to explicitly exploit the structure of the given nonsmoothness caused by the absolute value. For this purpose, the argument space is decomposed into polyhedra and then a sequence of quadratic optimization problems is solved on them. A given starting point determines the initial polyhedron and then optimality conditions help to identify a suitable neighboring polyhedron until a local minimizer is reached.

In [Kre23] this concept was extended to the Constrained Active Signature Method (CASM) for piecewise linear optimization problems with piecewise linear equality and inequality constraints. Here, an active set strategy is used to handle the constraints. For both ASM and CASM, necessary and sufficient optimality constraints, which can be verified in polynomial time, can be derived, see [GW18; Kre23]. Furthermore, in the same references, convergence in finitely many steps is proven. Optimality conditions for the more general nonlinear, nonsmooth case are considered in [HSS20].

In this section, first we will introduce some notations and then present the Constrained Active Signature Method for Abs-quadratic Problems, short QCASM, an algorithm derived from CASM to handle additional quadratic components in the objective function.

3.1 Notation and Definitions

The basis for developing the new optimization method is the so-called abs-linear form, which was introduced in [Gri13] and is defined as follows.

Definition 3.1 (abs-linear form, switching vector). *A continuous piecewise linear function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is in abs-linear form if $y \equiv f(x)$ is given by*

$$y = d + a^\top x + b^\top z, \quad (8a)$$

$$z = c + Zx + Mz + L|z|, \quad (8b)$$

with $x \in \mathbb{R}^n$ the argument vector, $z \in \mathbb{R}^s$ the vector of switching variables, called switching vector, and constants $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}^s$, $d \in \mathbb{R}$, $Z \in \mathbb{R}^{s \times n}$, $L, M \in \mathbb{R}^{s \times s}$, where the last two matrices are strictly lower triangular. Eq. (8b) is called switching system.

Here, $|z|$ denotes the component-wise absolute value of a vector z . In the context of optimization, one can set $d = 0$ without loss of generality. Furthermore, we assume that all components z_i , $1 \leq i < s$, are used as arguments of the absolute value for simplicity of notation. It follows with the identities

$$\min(x, y) = \frac{1}{2}(x + y - |x - y|) \quad \text{and} \quad \max(x, y) = \frac{1}{2}(x + y + |x - y|)$$

and the representation of piecewise linear function shown, e.g., in [Sch12, Proposition 2.2.2], that every piecewise linear function has a representation in an abs-linear form, see also [GW20, Lemma 2]. However, this representation might be not unique.

We denote by $\alpha(x) \equiv \{i \in \{1, \dots, s-1\} \mid z_i(x) = 0\}$ the set of indices of so-called active switching variables and by $P_\alpha \equiv (e_i^\top)_{i \in \alpha} \in \mathbb{R}^{|\alpha| \times s}$ the projection onto the active components of $z(x)$, where e_i is the i th unit vector of appropriate size. For each $x \in \mathbb{R}^n$ we define the signature vector by $\sigma(x) \equiv (\text{sgn}(z_1(x)), \dots, \text{sgn}(z_s(x))) \in \{-1, 0, 1\}^s$. The corresponding signature matrix is given by $\Sigma(x) = \text{diag}(\sigma(x))$. As in [Gri13], the signature vectors are used to define a polyhedral structure.

Definition 3.2 ((extended) signature domain). *For a fixed $\sigma \in \{-1, 0, 1\}^s$, define*

$$\mathcal{P}_\sigma \equiv \{x \in \mathbb{R}^n \mid \text{sgn}(z(x)) = \sigma\} \subset \overline{\mathcal{P}}_\sigma \equiv \{x \in \mathbb{R}^n \mid \Sigma z(x) = |z(x)|\}. \quad (9)$$

The set \mathcal{P}_σ is called signature domain and the set $\overline{\mathcal{P}}_\sigma$ extended signature domain.

As shown in [Gri+16, Proposition 2], the union of all possible extended signature domains equals \mathbb{R}^n . By definition, the extended signature domains do not have common interior points. Hence, they form a decomposition of \mathbb{R}^n . The points where the underlying piecewise linear function is not differentiable are located exclusively on the boundaries of the polyhedra. Motivated by the structure in \mathbb{R}^2 these boundaries are also called kinks, see for example [GW18; Kre23] for more details.

To incorporate also quadratic components in the objective function, we now extend the definition of abs-linear functions as follows.

Definition 3.3 (abs-quadratic form). *A continuous piecewise smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is in abs-quadratic form if $y \equiv f(x)$ is given by*

$$y = x^\top Q_1 x + x^\top Q_2 z + z^\top Q_3 z + a^\top x + b^\top z + d, \quad (10a)$$

$$z = c + Zx + Mz + L|z|, \quad (10b)$$

with $x \in \mathbb{R}^n$ the argument vector, $z \in \mathbb{R}^s$ the switching vector, and constants $Q_1 = Q_1^\top \in \mathbb{R}^{n \times n}$, $Q_2 \in \mathbb{R}^{n \times s}$, $Q_3 = Q_3^\top \in \mathbb{R}^{s \times s}$, $a \in \mathbb{R}^n$, $b, c \in \mathbb{R}^s$, $d \in \mathbb{R}$, $Z \in \mathbb{R}^{s \times n}$, $L, M \in \mathbb{R}^{s \times s}$, where the last two matrices are strictly lower triangular.

The chosen name is motivated by quadratic optimization problems, where the objective function is quadratic and the constraints are linear. In the further course of the paper, Eq. (10a) serves as the objective function of the considered optimization problem and Eq. (10b) represents one of the constraint sets. Note that in contrast to the piecewise linear case, not every piecewise quadratic function has an abs-quadratic form, as shown by the following example.

Example 3.4. *The function $f_1(x_1, x_2) = |x_1 x_2| = |x_1| |x_2|$ can be represented by the abs-quadratic form*

$$y = \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} \quad (11)$$

$$z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} |z_1| \\ |z_2| \\ |z_3| \\ |z_4| \end{pmatrix} \quad (12)$$

setting the remaining matrices and vectors to zero. However, the function $f_2(x_1, x_2) = |x_1 x_2 + 1|$ cannot be written in abs-quadratic form, since the switching equation does not allow a quadratic term.

3.2 Description of the Algorithm

First, we state the optimization problem for which we will present and analyze a solution algorithm in this paper, i.e., the *constrained abs-quadratic optimization problem* (CAQOP) given by

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & x^\top Q_1 x + x^\top Q_2 z + z^\top Q_3 z + a^\top x + b^\top z + d \\ \text{s.t.} \quad & 0 = g + Ax + Bz + C|z|, \\ & 0 \geq h + Dx + Ez + F|z|, \\ & z = c + Zx + Mz + L|z|, \end{aligned} \tag{CAQOP}$$

with an abs-quadratic target function and $g \in \mathbb{R}^m, h \in \mathbb{R}^p, A \in \mathbb{R}^{m \times n}, B, C \in \mathbb{R}^{m \times s}, D \in \mathbb{R}^{p \times n}$ and $E, F \in \mathbb{R}^{p \times s}$. For a detailed investigation of necessary and sufficient optimality conditions we refer to [GW16; HSS20; Kre23].

For later use, we introduce

$$H : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^s \rightarrow \mathbb{R}^p, \quad (x, z, |z|) \mapsto h + Dx + Ez + F|z|$$

as the function for the inequality constraints. Furthermore, we define $\omega = \text{sgn}(H(x, z, |z|))$ to identify the active inequalities and $\Omega = \text{diag}(\omega)$ as the projection onto these active inequality constraints.

The proposed optimization method is a feasible point approach. Hence, a feasible starting point x_0 is required that determines the initial polyhedron \mathcal{P}_σ with $\sigma = \sigma(x_0)$ to be examined. For such a fixed σ , we will now describe the steps of the optimization algorithm:

Computing a descent direction for given σ and ω Following the approaches presented in [GW18; Kre23], on each polyhedron defined by σ taking account the required signs of z , we obtain the optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^s} \quad & x^\top Q_1 x + x^\top Q_2 |\Sigma|z + z^\top |\Sigma|Q_3 |\Sigma|z + a^\top x + b^\top |\Sigma|z + d \\ \text{s.t.} \quad & 0 = g + Ax + B|\Sigma|z + C\Sigma z, \\ & 0 \geq h + Dx + E|\Sigma|z + F\Sigma z, \\ & 0 = |\Sigma|z - \tilde{c} - \tilde{Z}x, \\ & 0 \leq \Sigma z, \end{aligned} \tag{13}$$

with \tilde{Z} and \tilde{c} defined as

$$\tilde{Z} = (I_s - M - L\Sigma)^{-1}Z \quad \text{and} \quad \tilde{c} = (I_s - M - L\Sigma)^{-1}c. \tag{14}$$

Due to the fixed signature vector, this optimization problem is smooth, has a quadratic target function and linear constraints. When a solution exists, it can be solved with a standard QP method. However, we want to exploit the structure provided by the signature vector as additional feature. For this purpose, we apply standard KKT theory to Equation (13). With Lagrange multipliers $\delta \in \mathbb{R}^m, \nu \in \mathbb{R}^p, \lambda \in \mathbb{R}^s$ and $\mu \in \mathbb{R}^s$, we obtain the following necessary optimality conditions

$$0 = 2x^\top Q_1 + z^\top |\Sigma|Q_2^\top + a^\top + \delta^\top A + \nu^\top D - \lambda^\top \tilde{Z}, \tag{15a}$$

$$\begin{aligned} 0 = & x^\top Q_2 |\Sigma| + 2z^\top |\Sigma|Q_3 |\Sigma| + b^\top |\Sigma| \\ & + \delta^\top (B|\Sigma| + C\Sigma) + \nu^\top (E|\Sigma| + F\Sigma) + \lambda^\top |\Sigma| - \mu^\top \Sigma, \end{aligned} \tag{15b}$$

$$0 = g + Ax + B|\Sigma|z + C\Sigma z, \tag{15c}$$

$$0 \geq h + Dx + E|\Sigma|z + F\Sigma z, \tag{15d}$$

$$0 = |\Sigma|z - \tilde{c} - \tilde{Z}x, \tag{15e}$$

$$0 \leq \Sigma z, \quad 0 \leq \mu, \quad 0 = \mu^\top \Sigma z, \tag{15f}$$

$$0 \leq \nu, \quad 0 = \nu^\top (h + Dx + E|\Sigma|z + F\Sigma z). \tag{15g}$$

Multiplying Eq. (15b) by Σ from the right and using the sign condition for μ yields

$$0 \leq \mu^\top |\Sigma| = x^\top Q_2 \Sigma + 2z^\top |\Sigma|Q_3 \Sigma + b^\top \Sigma + \delta^\top (B\Sigma + C|\Sigma|) + \nu^\top (E\Sigma + F|\Sigma|) + \lambda^\top \Sigma. \tag{16}$$

Due to the complementarity condition $\mu^\top \Sigma z = 0$, this inequality must hold as an equation. Hence, it follows that

$$-b^\top \Sigma = x^\top Q_2 \Sigma + 2z^\top |\Sigma| Q_3 \Sigma + \delta^\top (B \Sigma + C |\Sigma|) + \nu^\top (E \Sigma + F |\Sigma|) + \lambda^\top \Sigma.$$

Thus with $\omega = \text{sgn}(H(x, z, |z|))$ and $\Omega = \text{diag}(\omega)$ denoting as before the projection onto the active inequality constraints, a solution x of the optimization task (13) together with the corresponding z and the Lagrange multipliers must satisfy the linear system

$$\begin{bmatrix} 2Q_1 & Q_2 |\Sigma| & -\tilde{Z}^\top & A^\top & D^\top \\ \Sigma Q_2^\top & 2\Sigma Q_3 |\Sigma| & \Sigma & \Sigma B^\top + |\Sigma| C^\top & \Sigma E^\top + |\Sigma| F^\top \\ \tilde{Z} & -|\Sigma| & 0 & 0 & 0 \\ A & B |\Sigma| + C \Sigma & 0 & 0 & 0 \\ \bar{\Omega} D & \bar{\Omega} (E |\Sigma| + F \Sigma) & 0 & 0 & \Omega \end{bmatrix} \begin{bmatrix} x \\ z \\ \lambda \\ \delta \\ \nu \end{bmatrix} = - \begin{bmatrix} a \\ \Sigma b \\ \tilde{c} \\ g \\ \Omega h \end{bmatrix}, \quad (17)$$

where $\bar{\Omega} = I_p - |\Omega|$ forces the active inequalities to vanish. The matrix Ω in the right lower corner ensures that ν is zero for the inactive inequality constraints. A solution of this linear system must exist as soon as the optimization task (13) has at least one solution. At the end of this section, we will discuss options to ensure the existence of such a solution. For the time being, we assume that we can compute a solution $(\hat{x}, \hat{z}, \lambda, \delta, \nu)$ of this linear system and define for the current iterate x and z

$$\Delta x := \hat{x} - x \quad \text{and} \quad \Delta z := \hat{z} - z$$

as directions towards the next iterate.

Computing a step size β The optimization problem (CAQOP) has only (piecewise) linear constraints. This situation was already analyzed for the CASM solver. Hence the step size strategy from CASM [Kre23] can be reused here. Nevertheless, we sketch the step calculation briefly for a complete picture of the algorithm.

For a solution (\hat{x}, \hat{z}) of Equation (17), we must check whether $\sigma(\hat{x}) = \sigma$ is still valid and that the inequality constraints of Eq. (15) still hold to ensure feasibility. For this purpose, we calculate two step sizes. The first one is the step length from the current iterate x in the direction Δx to a possible kink, i.e., a sign change in one component of z . It is denoted by β^z and defined as

$$\beta^z = \inf_{1 \leq l \leq s} \left\{ \beta_l^z \equiv \frac{-z_l}{\hat{z}_l - z_l} \mid (\hat{z}_l - z_l) \sigma_l < 0 \right\} \in (0, \infty]. \quad (18)$$

If $\beta^z < \infty$ the first index for which the minimum is attained is denoted by j^z . For $\beta^z < 1$, there exists a blocking kink such that we can not take the full step to the solution \hat{x} . For a given signature vector σ with $\sigma_i \neq 0$ and $x \in \mathcal{P}_\sigma$ it follows that $z_i(x) \neq 0$. Thus, $\beta^z > 0$ must hold.

The second step size is the step length from the current iterate x in the direction Δx to a possible inequality constraint $H_l(x, z, \Sigma z)$, $1 \leq l \leq p$, that becomes active. Similar to the computation of β^z , this step size β^H is given by

$$\beta^H = \inf_{1 \leq l \leq p} \left\{ \beta_l^H \equiv \frac{H_l}{H_l - \hat{H}_l} \mid (\hat{H}_l - H_l) \omega_l < 0 \right\} \in (0, \infty], \quad (19)$$

where $H \equiv H(x, z, \Sigma z)$, $\hat{H} \equiv H(\hat{x}, \hat{z}, \Sigma \hat{z})$ and l denotes the l th component of H and \hat{H} , respectively. We denote by j^H the smallest index for which the minimum is attained. If $\beta^H < 1$, there exists a blocking inequality constraint, i.e., the solution \hat{x} is not feasible. Therefore, the new iterate x^+ should be chosen such that the j^H th components of $H(x^+, z^+, \Sigma z^+)$ and $\omega(x^+)$ drop to zero in comparison to $H(x, z, \Sigma z)$ and ω , respectively. Setting $\omega_{j^H}^+ = 0$ changes the optimality system (15) and a new solution of system (17) has to be computed. If $\beta^H \leq \beta^z$ then we have $z_{j^H} \hat{z}_{j^H} \geq 0$ such that the iterate \hat{x} is still contained in \mathcal{P}_σ , i.e., $\sigma(\hat{x}) = \sigma$ is still valid. Since all active constraints are encoded in ω , one must have $\beta^H > 0$.

Finally, we determine the actual step size

$$\beta = \min\{\beta^z, \beta^H, 1\} \in (0, 1], \quad (20)$$

where the upper bound 1 on β ensures with the update

$$x^+ = (1 - \beta)x + \beta \hat{x} = x + \beta \Delta x$$

that the next iterate is feasible for the optimization task (13) and the current σ . As can be seen, the case $\beta < 1$ corresponds to the activation of a kink or an inequality constraint, respectively. Therefore, we will refer to this situation as a restriction of σ or ω , respectively. If $\beta = 1$, one has for the new iterate $x^+ = \hat{x}$ that $\sigma(x^+) = \sigma$ and $\omega(x^+) = \omega$. In this case, x^+ is called *signature stationary* since the two signature vectors are kept.

Algorithm 1 Quadratic Constrained Active Signature Method (QCASM)

Require: Feasible start point $x \in \mathbb{R}^n, n \in \mathbb{N}, s, m, p \in \mathbb{N} \cup \{0\}, a \in \mathbb{R}^n, b, c \in \mathbb{R}^s, Z \in \mathbb{R}^{s \times n}, L, M \in \mathbb{R}^{s \times s}$ strictly lower triangular, $Q_1 = Q_1^\top \in \mathbb{R}^{n \times n}, Q_2 \in \mathbb{R}^{n \times s}, Q_3 = Q_3^\top \in \mathbb{R}^{s \times s}, g \in \mathbb{R}^m, h \in \mathbb{R}^p, A \in \mathbb{R}^{m \times n}, B, C \in \mathbb{R}^{m \times s}, D \in \mathbb{R}^{p \times n}, E, F \in \mathbb{R}^{p \times s}, \beta = 0$

Set: $z := z(x)$ via Eq. (8b), $\sigma := \sigma(x)$ and $\omega := \omega(x)$

```
1: loop
2:   If possible compute  $(\hat{x}, \hat{z}, \lambda, \delta, \nu)$  by solving Eq. (17) else return ERROR
3:   Compute  $\beta^z$  via Eq. (18),  $\beta^H$  via Eq. (19) and  $\beta$  via Eq. (20)
4:   Set  $(x^+, z^+) = (1 - \beta)(x, z) + \beta(\hat{x}, \hat{z})$ 
5:   if  $\beta^H = \beta$  then Restrict  $\omega$  ▷ Add constraint
6:   if  $\beta^z = \beta$  then Restrict  $\sigma$  ▷ Add kink
7:   if  $\beta = 1$  then ▷  $x^+$  is feasible signature stationary
8:     if  $\nu \not\geq 0$  then
9:       Relax  $\omega$ , set  $\beta = 0$  ▷ Drop constraint
10:    else ▷  $x^+$  is feasible signature optimal
11:      if Eq. (21) holds true then
12:        Relax  $\sigma$ , set  $\beta = 0$  ▷ Drop kink
13:      else ▷  $x^+$  is local optimal
14:        return  $(x^+, z^+)$  ▷ Problem solved
15:    Set  $(x, z) = (x^+, z^+)$ 
```

Checking the optimality If x^+ is signature stationary on the current polyhedron \mathcal{P}_σ , it fulfills the necessary optimality condition of the optimization task (13) on \mathcal{P}_σ . Then, one has to check whether x^+ is a minimizer of (CAQOP). If this is the case the iteration stops. Otherwise, the optimization continues in one of the neighboring polyhedra $\mathcal{P}_{\tilde{\sigma}}$ with $\tilde{\sigma} \succ \sigma$ where

$$\tilde{\sigma} \succeq \sigma \quad \text{holds if} \quad \tilde{\sigma}_j \sigma_j \geq \sigma_j^2 \quad \text{for } j = 1, \dots, s.$$

Such a $\tilde{\sigma}$ can be decomposed into $\sigma + \gamma$ with $|\sigma|^\top |\gamma| = 0$. Replacing Σ in the optimality conditions (15) by the corresponding $\Sigma + \Gamma$ and using

$$z_{\tilde{\sigma}}(x) = z_{\sigma+\gamma}(x) = (I_s - M - L\Sigma - L\Gamma)^{-1}(c + Zx) = (I_s - \tilde{L}\Gamma)^{-1}(\tilde{c} + \tilde{Z}x)$$

it follows that most of the equations still hold for the current values of \hat{x}, \hat{z} and λ . Just Eq. (16) changes in that it has as many new nontrivial component as γ which are given by

$$\begin{aligned} 0 \leq \mu^\top |\Gamma| &= \hat{x}^\top Q_2 \Gamma + 2\hat{z}^\top |\Gamma| Q_3 \Gamma + b^\top \Gamma + \delta^\top (B\Gamma + C|\Gamma|) + \nu^\top (E\Gamma + F|\Gamma|) + \lambda^\top (I_s - \tilde{L}\Gamma)\Gamma \\ &= \left(\hat{x}^\top Q_2 + 2\hat{z}^\top |\Gamma| Q_3 + b^\top + \delta^\top B + \nu^\top E + \lambda^\top \right) \Gamma + (\delta^\top C + \nu^\top F - \lambda^\top \tilde{L})|\Gamma|. \end{aligned}$$

This condition is violated if and only if there exists at least one index k with $1 \leq k < s$ such that $\gamma \equiv e_k \operatorname{sgn}(b^\top + \delta^\top B + \nu^\top E + \lambda^\top)$ satisfies

$$0 > (\delta^\top C + \nu^\top F - \lambda^\top \tilde{L})e_k - \left| \hat{x}^\top Q_2 + 2\hat{z}^\top Q_3 b^\top + \delta^\top B + \nu^\top E + \lambda^\top \right| e_k \quad \text{and} \quad \sigma_k = 0. \quad (21)$$

This optimality condition can be verified as a matrix-vector product in polynomial time. For the unconstrained and abs-linear case a similar complexity result was shown in [GW18] and for the general case of abs-smooth and constrained problems in [HSS20].

The overall algorithm Combining all the steps described above, one obtains Algo. 1 as an extension of [KWG21, Algorithm 1] for the piecewise linear case. Once more, the resulting algorithm consists of three main parts: First, the computation of the search direction (cf. line 2 of Algo. 1). Second, computing the step size and in case of blocking kinks and/or inequality constraints restrict σ and/or ω , respectively (cf. line 3-6). Third, checking the optimality and relaxing kinks or constraints in case of nonoptimality (cf. line 7-14).

Finite convergence of QCASM For the piecewise linear case, convergence properties of (C)ASM have already been analyzed in detail in [GW18; Kre23]. The basic idea is that by decomposing \mathbb{R}^n there are only finitely many polyhedra and on one polyhedron a solution is found in finitely many steps. However, the essential difference lies in the quadratic parts of the objective function. For the piecewise

linear case considered in [GW18; Kre23] the resulting linear problems on each polyhedron were extended by a quadratic penalty term with a positive definite matrix Q to ensure the existence of a minimizer in each polyhedron. To derive the same convergence results for the optimization task (CAQOP), a similar requirement would be that the matrix

$$Q = \begin{pmatrix} Q_1 & \frac{1}{2}Q_2 \\ \frac{1}{2}Q_2^\top & Q_3 \end{pmatrix}$$

is positive definite. However, this is not necessarily the case for the examples considered in the earlier sections of this paper. On the other hand, it was shown in the Theorems 2.1, 2.2, and 2.4 that the target function is convex on the feasible domain. Therefore, we will now analyze the convergence of QCASM under more general convexity assumptions including also the case that the matrix Q is positive definite. For this purpose, first we define a regularity condition, which has already been introduced [GW19] for the unconstrained case and in [HSS20] for the constrained case. It corresponds to the well-known LICQ in the smooth case.

Definition 3.5 (LIKQ (constrained case)). *Let a constrained optimization problem of the form (CAQOP) and a signature vector $\sigma \in \{-1, 0, 1\}^s$ be given. Then the linear independent kink qualification (LIKQ) holds at a feasible point x of the corresponding optimization problem (13) if the active Jacobian*

$$\mathcal{J}_\sigma \equiv \begin{bmatrix} A + B|\Sigma|\tilde{Z} + C\Sigma\tilde{Z} \\ P_{\mathcal{I}}(D + E|\Sigma|\tilde{Z} + F\Sigma\tilde{Z}) \\ P_\alpha\tilde{Z} \end{bmatrix} \in \mathbb{R}^{(m+|\mathcal{I}|+|\alpha|) \times n} \quad (22)$$

has full row rank $m + |\mathcal{I}| + |\alpha|$. Here, $\mathcal{I} \equiv \mathcal{I}(x)$ collects the indices of the active inequality constraints at x and the projection onto the active components of $H(x)$ is defined as $P_{\mathcal{I}} \equiv (e_i^\top)_{i \in \mathcal{I}} \in \mathbb{R}^{|\mathcal{I}| \times p}$ with e_i denoting the i th unit vector of appropriate size.

Now, we present one of the main contributions of this paper.

Theorem 3.6 (Finite convergence of QCASM). *Let an optimization problem of the form (CAQOP) be given and suppose that LIKQ holds at every feasible point. Furthermore, assume that the target function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ of (CAQOP) is convex on the compact feasible set \mathcal{F} . Then, Algorithm 1 terminates for any feasible starting point $x \in \mathbb{R}^n$ after finitely many iterations at a global minimizer of (CAQOP).*

Proof. Since f is continuous and \mathcal{F} is compact, there exists a minimizer of the optimization task (CAQOP). For a fixed given σ , the same statement holds for the optimization task (13), such that one can always compute a solution of Eq. (17) and Algorithm 1 does not terminate with an error. Moreover, it is clear that if the current iterate x_i is not a minimizer of (CAQOP), then a descent direction must exist. Now, we show that such a descent direction is determined by Algorithm 1.

Since standard KKT theory was used to derive the system of equations (17), its solution is a stationary point for the currently active set of constraints, encoded by the matrix Ω , and the active switches, encoded by the matrix Σ . Since f is convex on the feasible set, such a stationary point is already a minimum for the restricted subproblem defined by the current Ω and Σ . If the solution of (17) yields a point \hat{x} that is not equal to the current iterate x , then $\Delta x = \hat{x} - x \neq 0$ defines a direction of descent. If $\hat{x} = x$, but x is not a minimizer of the original problem (CAQOP), then active inequality constraints and/or switches must be dropped. Since we can ensure in the next iterate that the step length β is larger than zero, a cycling can not happen. Hence, standard active set theory yields the convergence to a minimizer of the optimization task (13) in finitely many steps.

There are only finitely many polyhedra. Furthermore, if the signature vector σ is modified by Algo. 1, the value of the target function is consistently reduced in the next iterate. Hence, the signature vector can be modified only finitely many times leading to a finite convergence of the overall algorithm. ■

4 Numerical Validation

This section is organized into two subsections. The first subsection discusses a few possible data sources that have been used in the literature. This is included for the benefit of the community and aims at showing the reach of the problem and the importance of such datasets. We also note that most of these datasets are proprietary and are far from being published in open-source repositories. The second subsection presents the numerical performance of our proposed approach on three retail datasets. Two out of the three datasets correspond to real data and one refers to simulated data.

4.1 Data Sources

We divide this subsection further into three parts, where the first one elaborates on case studies that are proprietary, the second one on the datasets we use for this work, and the third one on other possible sources that can be used for the future.

4.1.1 Propriety Industry Data

We present two use-cases that specifically focus on problems pertaining to discounts and pricing.

- Assortment Optimization Case (2017): This work presents an inventory management problem, where the demand is a consequence of product prices. Here, promotions and markdowns follow a slightly different structure and apply not to the product pricing, but instead to their shelf assortment within the store (product sales is limited by an upper bound, which is a consequence of some pre-defined rules set by the store).

$$s_i^t = \min(d_i^t(p), \bar{s}_i^t), \quad d_i^t(p_i^t) = \max(a_i - b_i p_i^t, 0).$$

Note that \bar{s}_i^t is a constant, which is determined based on the various shelf widths and heights, specific to the store. Recent work [Hek+19] investigates this setting in detail based on real business data from a Turkish grocery store. However, the data has not been made publicly available.

- Markdown Optimization Case (2020/2021): Recent work [Bur+21] considers the promotions planning problem (discrete version) in spirit similar to our proposed formulation. However, the setting is slightly more complex, where the objective includes metrics that are not restricted to only revenue. This data is from a European retail outlet (details also anonymized) and is not publicly available.

4.1.2 Open Source Repositories

While there are multiple datasets available from the standpoint of retail, e.g., market basket data, product attribute data, inventory data, and geographical variation data, very few actually correspond to historical price-sales entries. Of these, we describe two crucial and widely used datasets and one simulated dataset as follows.

- Tech-Gadget Retailer (Cohen’s dataset): This dataset is anonymized and refers to real data from a US based retailer [Coh+22]. It comprises 44 SKUs and 4400+ entries, with individual columns for price and sales. This dataset has been used as one of the standards for demand forecasts, e.g., by [CZJ19; DLS22]. The time stamps for the records range from the years of 2016 to 2018 for 98 weeks.
- Online Retail: This refers to data from the UCI repository [CSG12] and contains all the transactions between Jan. 2010 and Sep. 2011 for a real UK based online retail company (details similarly anonymized). The customers are wholesalers and the products/SKUs are “gifts” (occasion specific). The dataset has a few additional indicators like geography. It contains more than 2500 SKUs and roughly 50,000 records (transactions).
- Logit: Here, we simulate data based on previous work on logit models [EC23]. While the logit model is well established and possibly the most widely used, most of these models also contain product attributes or competition related additions. The model takes the form: $d(p) = \{1 + \exp\{-\psi(p)\}\}^{-1}$. We use the same parameters from [EC23], adding minor randomization with respect to the products. Our data is generated for 50 products and 50 weeks.

4.1.3 Other References

Some datasets from kaggle [Che; SB21] and beyond [Ske+21] also contains the necessary price-sales entries. However, they comprise information on product attributes in addition to prices. These datasets have been mainly analyzed in literature from the standpoint of clustering or association rule mining. Demand forecasting in these cases has also been studied, but over these comprehensive input features. These datasets hence do not ideally suit our case study.

4.2 Numerical Results

We present three sets of results in this discussion. The numbers denoting the revenue are expressed in the respective currency units (US-Dollars or Pounds). First, we analyze the progressive performance of the QCASM algorithm for the different datasets. Second, we discuss the accuracy of the three different piecewise linear functions on each of these datasets. Lastly, we compare the results of our optimization algorithm with standard base price solutions and randomly generated prices. All our computations were done on a quadcore Linux machine (Ubuntu 15.0) with 2.7Ghz speed and 32 GB RAM. We consider a simple version of the revenue maximization problem, which is continuous and further ignore inter-item constraints. Prices are bounded from above and below, where upper bounds indicate the base price. The lower bounds and the interior values of price indicate discounts. The price bounds are set to be the lowest and highest values of the respective products prices from historical data. The initial inventory and

Iteration	Revenue (in multiples of 100,000)					
	Product 4	Product 5	Product 7	Product 8	Product 12	Total
1	0.07529	0.04287	0.11084	1.14629	0.56539	24.4017
10	0.07620	0.04380	0.11968	1.16121	0.56647	25.4976
20	0.07734	0.04496	0.13073	1.17985	0.56783	26.8673
50	0.08075	0.04844	0.16388	1.23578	0.57191	30.9767
70	0.08121	0.04891	0.18598	1.27306	0.57191	32.3248
80	0.08121	0.04891	0.19703	1.29170	0.57191	32.8250
100	0.08121	0.04891	0.21913	1.32899	0.57191	33.8253

Table 1: Iterative progress of QCASM for Cohen’s problem

replenishment levels are set to be proportionate to the total historical sales. All the datasets considered in our study and the related implementation details are available online¹.

Iterative Performance: We begin our analysis with Cohen’s dataset. For this dataset, the sales data obviously appeared to follow a non-concave trajectory and hence the choice of $\max(.,.)$ is ideal from both practicality and accuracy standpoints. Given the 44 SKUs, QCASM was used to solve both nonsmooth optimization problems, regression and revenue maximization. The termination criterion for both cases was 200 iterations of the active signature method or attainment of optimal error tolerance levels. Table 1 shows the revenue of different products at different stages of the QCASM algorithm. The estimation problem is relatively simple in that it consumes a significantly lesser number of iterations. Hence, we ignore reporting the related iterative performance. Here, coefficients a and b refer to the values from the fit $s_i^t \approx \max(a_i^0 - b_i^0 p, a_i^1 - b_i^1 p_i^t)$. It can be observed that convergence is reached in very few iterations for some products. This is very much expected since the variability in price and demand levels is huge across the product portfolio. In the case of nearly one-third of the products, the number of iterations did not exceed 10. For the UCI dataset, a similar iterative behavior is reported in table 2. Here, coefficients a and b are fitted with respect to the curve $s_i^t = \max(a_i - b_i p_i^t, 0)$ and the second problem (revenue maximization) is solved using the same expression. For this case, the sales data had a large fraction of “0” values and product returns, and the choice of $\max(., 0)$ is the most practical. The total revenue is reported in the last column. Similar results are reported in table 3 for another set of products. For

Iteration	Revenue									
	P-377	P-780	P-756	P-263	P-599	P-1060	P-441	P-686	P-324	Total
1	257.92	1939.60	132.60	1536.60	1234.48	515.84	792.48	1012.94	3307.20	5857376.16
10	425.60	2726.41	326.01	3119.66	2616.66	875.51	1056.24	1903.86	4852.65	7692706.11
20	635.20	3709.92	567.77	5098.49	4344.38	1325.10	1385.94	3017.51	6784.47	9987238.74
50	1264.01	6660.46	1293.04	11034.96	9527.54	2673.87	2375.05	6358.47	12579.92	16803284.78
70	1347.85	7053.87	1389.75	11826.49	10218.63	2853.70	2506.93	6803.94	13352.65	17766752.17
80	1347.85	7053.87	1389.75	11826.49	10218.63	2853.70	2506.93	6803.94	13352.65	17798875.94
100	1347.85	7053.87	1389.75	11826.49	10218.63	2853.70	2506.93	6803.94	13352.65	17824970.72
150	1347.85	7053.87	1389.75	11826.49	10218.63	2853.70	2506.93	6803.94	13352.65	17839027.50

Table 2: Iterative progress of QCASM for UCI’s problem

Iteration	Revenue									
	P-1169	P-1316	P-1316	P-1476	P-1145	P-1913	P-1541	P-1404	P-1433	P-1201
1	187.13	1505.40	1505.40	1491.85	1325.48	895.44	388.96	1549.08	1677.52	309.40
10	555.90	4884.62	4884.62	3248.09	3816.33	4273.68	1650.53	2033.86	2001.66	1068.77
20	1016.87	9108.65	9108.65	5443.38	6929.90	8496.49	3227.48	2639.84	2406.83	2017.98
50	2399.77	21780.72	21780.72	12029.27	16270.60	21164.91	7958.35	4457.78	3622.35	4865.62
70	2584.16	23470.33	23470.33	12907.38	17516.03	22854.03	8589.13	4700.18	3784.42	5245.30
80	2584.16	23470.33	23470.33	12907.38	17516.03	22854.03	8589.13	4700.18	3784.42	5245.30
100	2584.16	23470.33	23470.33	12907.38	17516.03	22854.03	8589.13	4700.18	3784.42	5245.30
150	2584.16	23470.33	23470.33	12907.38	17516.03	22854.03	8589.13	4700.18	3784.42	5245.30

Table 3: Iterative progress of QCASM for UCI’s problem (second set)

the logit dataset, a similar iterative behavior is reported in table 4. Here, coefficients a and b are fitted with respect to the curve $s_i^t = \min(a_i^0 - b_i^0 p_i^t, a_i - b_i p_i^t)$ and the second problem (revenue maximization)

¹Please follow this link for all details: <https://scm.cms.hu-berlin.de/aswinkannan1987/retail-code-base>.

is solved using the same expression. This solves the issue of concavity with respect to maximization of revenue and as we will see later also provides the best possible fit. The choice of the min type PLF for

Iteration	Revenue - in multiples of 100,000								
	P-38	P-8	P-11	P-36	P-24	P-21	P-25	P-33	Total
1	9.5455	13.6753	3.7833	9.9388	7.7593	8.4056	9.2360	7.5832	342.6827
10	9.9612	14.5872	5.6423	11.2419	9.7020	9.3681	10.1433	9.3743	428.4903
20	10.4809	15.7272	7.9662	12.8707	12.1304	10.5713	11.2774	11.6132	535.7499
50	12.0400	19.1471	14.9376	17.7572	19.4154	14.1807	14.6798	18.3299	857.5285
70	13.0794	21.4270	15.8672	18.4088	20.3867	16.5871	16.9481	19.2254	919.1238
80	13.5991	22.5669	15.8672	18.4088	20.3867	17.7902	18.0822	19.2254	930.8060
100	14.6385	24.8468	15.8672	18.4088	20.3867	20.1965	20.3505	19.2254	954.1704
150	14.9504	25.5308	15.8672	18.4088	20.3867	20.9184	21.0310	19.2254	961.1797

Table 4: Iterative progress of QCASM for the Logit problem

the logit case can be observed to be quite straightforward.

Comparison of models: In synopsis, the three piecewise linear models give a good estimate on the nature of demand for each of our datasets. If a min type function is more accurate, then demand can be possibly more concave. Table 5 shows the accuracy error (mean absolute error) for all the different possible scenarios. It follows that the $\max(.,.)$ fits the Cohen problem finer, and $\min(.,.)$ suits the concave demand problem.

Dataset	Mean Absolute Error (Scaled)		
	Model - $\max(.,0)$	Model - $\max(.,.)$	Model - $\min(.,.)$
Cohen	46.4562	42.7222	47.8897
UCI	19.9365	6.1840	8.6981
Concave-Demand	5.6640	5.6637	0.7258

Table 5: Comparison of Different Piecewise Linear Functions

Solution Quality: For the revenue maximization problem, we compare our results with three standard (and rudimentary) settings. One is setting the prices throughout at the base level (upper bound). The second setup refers to a randomly generated price vector between the lower and upper bounds, while the third one refers to the mid point between the lower and upper price bounds. It can be easily noticed from table 6 that there is a significant improvement in running our algorithm QCASM on the actual nonsmooth optimization problem. At this point, we do not intend to compare the performance of our algorithms against other smooth reformulations or other quadratic optimization methods. We just demonstrate as a proof of concept that the nonsmooth problem can be solved efficiently by the QCASM solver and the scheme is also theoretically convergent when demand functions follow some structure on concavity.

Dataset	Revenue			
	Base price	Random price	Midpoint price	QCASM
Cohen	2440173.82	2970553.95	2846046.30	4774583.20
UCI	5857376.16	15807897.04	12736676.95	17841259.92
Concave-Demand	34268272.32	88747048.32	76545493.35	96117974.13

Table 6: Comparison of Different Piecewise Linear Functions

5 Conclusion

This work considered nonsmooth quadratic optimization problems in retail revenue management. A continuous variant of the multi-period problem with inventory constraints and price bounds was analyzed and some preliminary statements on convexity were stated. The problem also required the estimation of piecewise linear functions that fit the historical data on price-sales leading to a similar optimization problem. A convergent active signature method that deploys the structure of the nonsmoothness was proposed for both problems in question. For a proof of concept, three practical datasets were studied. While the first two cases originate from real retail data, the third one was simulated from the widely used logit model. Preliminary results demonstrated a good behavior in iterative performance and revenue

improvement in comparison to other trivially chosen points. Future work aims to expand on deploying active signature methods for generalizations of our problem involving discrete prices and product design attributes.

Acknowledgment

The research was funded partly by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy — The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID:390685689).

References

- [All38] Roy G. D. Allen. **Mathematical analysis for economists**. English. Macmillan London, 1938.
- [Bur+21] Vibhati Burman et al. “Markdown Pricing For a Large Scale Retailer”. In: **2021 IEEE International Conference on Big Data (Big Data)**. 2021, pp. 1817–1826. DOI: 10.1109/BigData52589.2021.9671641.
- [CCH03] Wei-yu K. Chiang, Dilip Chhajed, and James Hess. “Direct Marketing, Indirect Profits: A Strategic Analysis of Dual-Channel Supply-Chain Design”. In: **Management Science** 49 (Jan. 2003), pp. 1–20. DOI: 10.1287/mnsc.49.1.1.12749.
- [CDH90] Choi S. Chan, Wayne S. Desarbo, and Patrick T. Harker. “Product Positioning under Price Competition”. In: **Management Science** 36.2 (1990), pp. 175–199. ISSN: 00251909, 15265501. URL: <http://www.jstor.org/stable/2661454>.
- [CG05] Daniel Corsten and Thomas Gruen. “On Shelf Availability: An Examination of the Extent, the Causes, and the Efforts to Address Retail Out-of-Stocks”. In: Jan. 2005, pp. 131–149. ISBN: 3-540-22611-7. DOI: 10.1007/3-540-27059-0_9.
- [Che] Daqing Chen. **Online retail dataset**. <https://www.kaggle.com/datasets/lakshmi25npathi/online-retail-dataset/metadata>.
- [Che+16] Xin Chen et al. “Dynamic Stochastic Inventory Management with Reference Price Effects”. In: **Operations Research** 64.6 (2016), pp. 1529–1536. DOI: 10.1287/opre.2016.1524.
- [Chi01] Pradeep K. Chintagunta. “Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data”. In: **Marketing Science** 20.4 (2001), pp. 442–456. DOI: 10.1287/mksc.20.4.442.9751.
- [Coh15] Maxime C. Cohen. “Pricing for retail, social networks and green technologies”. PhD thesis. Massachusetts Institute of Technology, Jan. 2015.
- [Coh+17] Maxime C. Cohen et al. “The Impact of Linear Optimization on Promotion Planning”. In: **Operations Research** 65.2 (2017), pp. 446–468. DOI: 10.1287/opre.2016.1573.
- [Coh+22] Maxime C. Cohen et al. **Demand Prediction in Retail A Practical Guide to Leverage Data and Predictive Analytics**. Springer Series in Supply Chain Management, 14. 2022.
- [CPP16] Maxime C. Cohen, Georgia Perakis, and Robert S. Pindyck. “Pricing with Limited Knowledge of Demand”. In: **Proceedings of the 2016 ACM Conference on Economics and Computation**. EC ’16. Association for Computing Machinery, 2016, p. 657. DOI: 10.1145/2940716.2940734.
- [CRS06] Youhua (Frank) Chen, Saibal Ray, and Yuyue Song. “Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales”. In: **Naval Research Logistics (NRL)** 53.2 (2006), pp. 117–136. DOI: <https://doi.org/10.1002/nav.20127>.
- [CSG12] Daqing Chen, Sai Laing Sain, and Kun Guo. “Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining”. In: **Journal of Database Marketing and Customer Strategy Management** 19.3 (Aug. 2012), pp. 197–208. DOI: 10.1057/dbm.2012.17.

- [CSL04] Xin Chen and David Simchi-Levi. “Coordinating Inventory Control and Pricing Strategies with Random Demand and Fixed Ordering Cost: The Finite Horizon Case”. In: **Operations Research** 52.6 (2004), pp. 887–896. URL: <http://www.jstor.org/stable/30036639>.
- [CZJ19] Maxime C. Cohen, Renyu (Philip) Zhang, and Kevin Jiao. “Data Aggregation and Demand Prediction”. In: **ERN: Statistical Decision Theory; Operations Research (Topic)** (2019).
- [DHP10] Maarten Dossche, Freddy Heylen, and Dirk Van den Poel. “The Kinked Demand Curve and Price Rigidity: Evidence from Scanner Data*”. In: **The Scandinavian Journal of Economics** 112.4 (2010), pp. 723–752. DOI: <https://doi.org/10.1111/j.1467-9442.2010.01621.x>.
- [DLS22] Yiting Deng, Yuexing Li, and Jing-Sheng Jeannette Song. “A Unified Parsimonious Model for Structural Demand Estimation Accounting for Stockout and Substitution”. In: **SSRN** (2022).
- [EC23] Moshe Eben-Chaime. “Demand Models, Revenue Curves and Profit”. In: **Athens Journal of Business and Economics** 9.2 (2023), pp. 221–230. DOI: 10.1287/msom.2015.0561.
- [FGL03] Gadi Fibich, Arieh Gavious, and Oded Lowengart. “Explicit Solutions of Optimization Models and Differential Games with Nonsmooth (Asymmetric) Reference-Price Effects”. In: **Operations Research** 51.5 (2003), pp. 721–734. DOI: 10.1287/opre.51.5.721.16758.
- [Gre95] Eric A. Greenleaf. “The Impact of Reference Price Effects on the Profitability of Price Promotions”. In: **Marketing Science** 14.1 (1995), pp. 82–104. ISSN: 07322399, 1526548X. URL: <http://www.jstor.org/stable/184209>.
- [Gri13] Andreas Griewank. “On stable piecewise linearization and generalized algorithmic differentiation”. In: **Optimization Methods and Software** 28.6 (Apr. 2013), 1139–1178. DOI: 10.1080/10556788.2013.796683.
- [Gri+16] Andreas Griewank et al. “On Lipschitz optimization based on gray-box piecewise linearization”. In: **Mathematical Programming** 158 (2016), pp. 383–415.
- [GW16] Andreas Griewank and Andrea Walther. “First- and second-order optimality conditions for piecewise smooth objective functions”. In: **Optimization Methods and Software** 31.5 (2016), pp. 904–930. DOI: 10.1080/10556788.2016.1189549.
- [GW18] Andreas Griewank and Andrea Walther. “Finite convergence of an active signature method to local minima of piecewise linear functions”. In: **Optimization Methods and Software** (Dec. 2018), pp. 1–21. DOI: 10.1080/10556788.2018.1546856.
- [GW19] Andreas Griewank and Andrea Walther. “Relaxing kink qualifications and proving convergence rates in piecewise smooth optimization”. In: **SIAM J. Optim.** 29.1 (2019), pp. 262–289. ISSN: 1052-6234. DOI: 10.1137/17M1157623. URL: <https://doi.org/10.1137/17M1157623>.
- [GW20] Andreas Griewank and Andrea Walther. “Beyond the Oracle: Opportunities of Piecewise Differentiation”. In: **Numerical Nonsmooth Optimization: State of the Art Algorithms**. Ed. by Adil M. Bagirov et al. Cham: Springer International Publishing, 2020, pp. 331–361. DOI: 10.1007/978-3-030-34910-3_10.
- [Hek+19] Mustafa Hekimoglu et al. “Assortment optimization with log-linear demand: Application at a Turkish grocery store”. In: **Journal of Retailing and Consumer Services** 50 (Sept. 2019), pp. 199–214. DOI: 10.1016/j.jretconser.2019.04.007.
- [HLP13] Jian Huang, Mingming Leng, and Mahmut Parlar. “Demand Functions in Decision Modeling: A Comprehensive Survey and Research Directions”. In: **Decision Sciences** 44.3 (2013), pp. 557–609. DOI: <https://doi.org/10.1111/deci.12021>.
- [HSS20] Lisa C. Hegerhorst-Schultchen and Marc C. Steinbach. “On first and second order optimality conditions for abs-Normal NLP”. In: **Optimization** 69.12 (2020), pp. 2629–2656. DOI: 10.1080/02331934.2019.1626386.

- [Hua+21] Junhao Hua et al. “Markdowns in E-Commerce Fresh Retail: A Counterfactual Prediction and Multi-Period Optimization Approach”. In: **CoRR** abs/2105.08313 (2021). URL: <https://arxiv.org/abs/2105.08313>.
- [KC14] Timo Kunz and Sven Crone. “Demand models for the static retail price optimization problem – A Revenue Management perspective”. In: **OpenAccess Series in Informatics** 37 (May 2014). DOI: 10.4230/OASICS.SCOR.2014.101.
- [KC62] Samuel Karlin and Charles R. Carr. “Prices and Optimal Inventory Policy”. In: **Studies in Applied Probability and Management Science**. Ed. by Kenneth J. Arrow, Samuel Karlin, and Herbert Scarf. Stanford (Calif.): Stanford University Press, 1962. Chap. 10, pp. 159–172.
- [KP20] Aswin Kannan and Kiran Panchangam. **Computerized promotion and markdown price scheduling**. US Patent 10,776,803. Sept. 2020.
- [KPW20] Aswin Kannan, Kiran Panchangam, and Su-Ming Wu. **Computerized promotion and markdown price scheduling**. US Patent 10,528,903. Jan. 2020.
- [Kre23] Timo Kreimeier. “Solving Constrained Piecewise Linear Optimization Problems by Exploiting the Abs-Linear Approach”. PhD thesis. Humboldt-Universität zu Berlin, 2023.
- [KWG21] Timo Kreimeier, Andrea Walther, and Andreas Griewank. **An active signature method for constrained abs-linear minimization**. Tech. rep. Humboldt-Universität zu Berlin, 2021. URL: https://opus4.kobv.de/opus4-trr154/files/474/main_casm_2023.pdf.
- [Li+22] Hanwei Li et al. “Large-Scale Price Optimization for an Online Fashion Retailer”. In: **Innovative Technology at the Interface of Finance and Operations: Volume II**. Springer International Publishing, 2022, pp. 191–224. DOI: 10.1007/978-3-030-81945-3_8.
- [LL03] Amy Hing Ling Lau and Hon-Shiang Lau. “Effects of a demand-curve’s shape on the optimal solutions of a multi-echelon inventory/pricing model”. In: **European Journal of Operational Research** 147.3 (2003), pp. 530–548. ISSN: 0377-2217. DOI: 10.1016/S0377-2217(02)00291-6.
- [MF17] Shaohui Ma and Robert Fildes. “A retail store SKU promotions optimization model for category multi-period profit maximization”. In: **European Journal of Operational Research** 260.2 (2017), pp. 680–692. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2016.12.032.
- [Mil59] Edwin S. Mills. “Uncertainty and Price Theory”. In: **The Quarterly Journal of Economics** 73.1 (1959), pp. 116–130. ISSN: 00335533, 15314650. URL: <http://www.jstor.org/stable/1883828>.
- [Mit+22] Arnab Mitra et al. “A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach”. In: **SN Operations Research Forum** 3.4 (Dec. 2022), pp. 1–22. DOI: 10.1007/s43069-022-00166-.
- [NC74] Masao Nakanishi and Lee G. Cooper. “Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach”. In: **Journal of Marketing Research** 11.3 (1974), pp. 303–311. ISSN: 00222437. URL: <http://www.jstor.org/stable/3151146>.
- [RLS05] Saibal Ray, Shanling Li, and Yuyue Song. “Tailored Supply Chain Decision Making under Price-Sensitive Stochastic Demand and Delivery Uncertainty”. In: **Management Science** 51.12 (2005), pp. 1873–1891. ISSN: 00251909, 15265501. URL: <http://www.jstor.org/stable/20110474>.
- [SB21] Brian Seaman and John Bowman. “Applicability of the M5 to Forecasting at Walmart”. In: **International Journal of Forecasting** (2021). ISSN: 0169-2070. DOI: <https://doi.org/10.1016/j.ijforecast.2021.06.002>.
- [Sch12] Stefan Scholtes. **Introduction to Piecewise Differentiable Equations**. Springer-Briefs in Optimization. New York, NY: Springer New York, 2012.

- [Ske+21] Geri Skenderi et al. **Well Googled is Half Done: Multimodal Forecasting of New Fashion Product Sales with Image-based Google Trends**. 2021. DOI: 10.48550/ARXIV.2109.09824.
- [SRL08] Yuyue Song, Saibal Ray, and Shanling Li. “Structural Properties of Buyback Contracts for Price-Setting Newsvendors”. In: **Manufacturing & Service Operations Management** 10.1 (2008), pp. 1–18. DOI: 10.1287/msom.1060.0140.
- [SS10] Shivaram Subramanian and Hanif Sherali. “A fractional programming approach for retail category price optimization”. In: **J. Global Optimization** 48 (Oct. 2010), pp. 263–277. DOI: 10.1007/s10898-009-9491-2.
- [TR19] Jean Tirole and Patrick Rey. “Price Caps as Welfare-Enhancing Coopetition”. In: **Journal of Political Economy** 127.6 (2019), pp. 3018–3069. DOI: 10.1086/702014.
- [YGT14] Ghasemy R. Yaghin, S.M.T. Fatemi Ghomi, and S.A. Torabi. “Enhanced joint pricing and lotsizing problem in a two-echelon supply chain with logit demand function”. In: **International Journal of Production Research** 52.17 (2014), pp. 4967–4983. DOI: 10.1080/00207543.2014.885665.